The Rising Bar to Entrepreneurship: Evidence from France

Maxime Bonelli [*] Jack Liebersohn[†] Victor Lyonnet[‡]

12[th] May, 2021

Abstract

Research has found evidence for a world-wide decline in the number of new firms, but less is known about how the quality of new firms has changed. Using representative data from a large-scale French survey of entrepreneurs, we document that the ex-ante ability of entrepreneurs increased from 1998-2018. To measure quality, we estimate which characteristics predict ex post success in the 1998 cohort and predict the likelihood of growth in subsequent years. A variety of approaches confirms rising ex ante quality. We show that quality increases are greater in commuting zones and industries with a greater decline in entrepreneurship, and a shift-share approach confirms a causal relationship between fewer firms and higher ability. Our results suggest that low-ability entrepreneurs are the least likely to continue, increasing the average quality over time.

[*]HEC Paris. Email: maxime.bonelli@hec.edu
[†]UC Irvine. Email: cjlieber@uci.edu
[‡]Ohio State University. Email: lyonnet.1@osu.edu

1 Introduction

Evidence from the United States and other developed countries has shown that the number of entrepreneurs has fallen since at least the 1990s (see, e.g., Decker et al., 2014, 2016a,b; Akcigit and Ates, 2021). Although new firms are responsible for the majority of new employment, a small number of successful firms make up a disproportionate amount of new job growth (Decker et al., 2016b). Given the importance of ex-ante entrepreneur characteristics in explaining the differences in outcomes across startups (Pugsley, Sedlacek, and Sterk, 2019), it is important to understand not only how the number of firms is changing, but how startup quality has evolved over time.

This paper measures how the quality of entrepreneurs has changed in France from 1998-2018 and provides evidence for the source of the changes. Measuring entrepreneur quality is difficult because ex post success depends on ex ante entrepreneur characteristics as well as economic trends and unforeseen events. The existing evidence on entrepreneurship mostly focuses on ex post measures of success, which are a result of entrepreneurs' traits and abilities as well as the economic conditions they face. This means that it is hard to know, for example, whether the decline in firm growth is because there are fewer "high-growth" entrepreneurs — reflecting a change in abilities — or because success is harder than it used to be. For example, the same conditions which have reduced the number of entrepreneurs may have made success more difficult, even if the quality of entrepreneurs has improved.

To circumvent these issues, this paper proposes a new methodology to measure variation in entrepreneurs' ability in a large sample of new French firms. Using data from the 1998 cohort of French entrepreneurs, we show success is partially predictable from entrepreneurs' work experience and demographic variables. After using a LASSO procedure to select predictive variables, we fit linear regressions of success on characteristics in the 1998 cohort, and calculate the fitted values for each later cohort. The result is an index of "predicted success" at the firm level that is based entirely on entrepreneurs' characteristics — a measure of ex ante entrepreneur ability. We use this measure to document trends in entrepreneur ability across regions and over time.

Several new facts emerge from our measures of predicted success. First, different characteristics predict firm survival, and growth conditional on survival. This means that the types of entrepreneurs whose firms are likely to survive for several years are not necessarily the same kinds whose firms are likely to hire many employees or to be very profitable. For example, entrepreneur age predicts survival, whereas previous executive experience predicts growth. The difference between "transformational" and "subsistence" entrepreneurs is well-understood in the literature (Schoar, 2010), but to the best of our knowledge, there have not been previous attempts to measure the relative prevalence of both types of entrepreneurs in the economy.

Second, there is substantial time variation in the characteristics of new entrepreneurs. Our data includes five cohorts sampled from 1998-2014, and over this time, there is an increase in the average predicted growth of new entrepreneurs. This means that, even as startups are a shrinking share of the economic pie in France, the entrepreneurs that do survive are likely to create high-growth firms. At the same time, there is a decrease in the predicted survival probability of new firms. Over all, entrepreneurs are increasingly drawn from a pool of individuals who create firms with high growth rates, but who are likely to have high failure rates.

These findings are made possible by the use of a large-scale representative survey of the population of French entrepreneurs. It is unique for several reasons: Its large sample size, high response rate and representativeness; its links to other administrative datasets, leading to a panel structure; and its comprehensive coverage of industries and geographies. The survey is conducted every four years from 1998 to 2014. We combine it with the corporate tax files and employment records available for all firms. Unlike most of the existing literature, our data allow us to test different measures of success and see whether different entrepreneur characteristics are predictive of different measures of success. Our preferred success measures are the employment level, and firm value-added, all measured at age 5.[1] Our data is very rich, allowing us to select among 48 entrepreneur characteristics (32 numerical variables and 16 categorical variables) and their interaction terms. We therefore have many potential predictors of which only a few might be important for predicting entrepreneurial success.

The survey includes questions on entrepreneurs' background, such as their education, employment history, and demographics; as well as questions about the choices they make as entrepreneurs, such as their financing and founding teams. We focus on variables measuring demographics and employment history, since these are related to the qualities of the entrepreneurs themselves, rather than the immediate economic conditions that they find themselves.

In the first part of the paper, we establish stylized facts about changes in entrepreneurship in France. As in the United States, the number of high-growth new French firms has fallen, as has the fraction of employment in new firms. One difference to the United States is that the overall number

[1] We code observations of exiting firms as "zeros" to account for both the intensive (magnitude) and extensive (survival) margins of success. Using other success measures at different ages, we find our qualitative results are not significantly affected.

of new firms has increased, which is explained by a growing number of sole proprietorships.

Next, we use the survey to ask which *ex-ante* entrepreneur characteristics are predictive of success conditional on entry. We use different non-parametric machine learning approaches to identify which variables best predict success. While traditional regressions are typically designed for estimating structural parameters and drawing causal inferences, machine learning algorithms are substantially better at making predictions.[2] In particular, the Least Absolute Shrinkage and Selection Operator (LASSO) method is well suited for variable prediction in our context. The penalty function in the LASSO results in a sparse estimator with many coefficients set exactly to zero, so that the LASSO estimator may be used for variable selection by simply selecting the variables with nonzero estimated coefficients. The LASSO method addresses heteroskedasticity, clustering, and non-normality in model errors (Belloni, Chernozhukov, and Hansen, 2014a). We alleviate the LASSO biased estimates by employing the Post-LASSO estimator that selects variables and then estimate the coefficients on these variables via ordinary least squares regression using only these variables (Belloni, Chernozhukov, and Hansen, 2014a,b). We check the robustness of our findings by implementing other machine-learning algorithms to select which entrepreneur characteristics best predict success (RIDGE, ELASTIC NET). Our favored approach is LASSO because it selects a restricted set of variables that best predict success.

Our data allow us to take a step towards a better understanding of the enormous heterogeneity across entrepreneurs. Some of this heterogeneity may reflect differences in aspirations and abilities across entrepreneurs. We estimate the importance of a wide array of entrepreneur characteristics for their startup's success. We find that the variables that best predict success are variables related to previous work experience, and a few demographic variables. Although we have rich education data available, it is not as predictive of success. We also find that similar results hold when we predict success using just the variables themselves as when we fit the LASSO using the full set of interaction terms. For ease of interpretability, we therefore rely on the smaller set of predictive variables.

Our key assumption is that the variables we select are determined before their firms are established. This allows us to interpret our quality index as measuring changing entrepreneur characteristics rather than being a direct proxy for changes in the business environment. As an example of this, we do not include variables measuring how new firms are financed, since this is a result of con-

[2] Machine learning algorithms are designed to maximize out of sample predictive accuracy by avoiding overfitting and by not being constrained by specific parametric assumptions or restrained in the number of covariates. See Athey and Imbens (2017) and Mullainathan and Spiess (2017).

temporaneous financial conditions, rather than the characteristics of the entrepreneurs themselves.

Importantly, we do not interpret the variables we select as "causing" greater success – rather, they are a proxy for underlying characteristics that are associated with success conditional on the decision to start a firm. For example, we find that entrepreneurs who have never started a firm before are more likely to survive until year 5. We do not think that this means experience with a startup *causes* failure; rather, this variable is a proxy for risk attitude, willingness to experiment, or the types of firms that these entrepreneurs start. Our interpretation is supported by the fact that, conditional on survival, first-time entrepreneurs' firms are smaller after five years than firms started by serial entrepreneurs.

Finally, we link changes in startup quality related to the decline in entrepreneurship. We run cross-sectional regressions to estimate how changes in firm entry at the sector level are related to changing entrepreneur quality. We find that sectors with greater increases in value-add among established firms are exactly the sectors where there is more entry. These are also the sectors that have experienced the smallest increase in entrepreneur "quality", as measured by predicted employment. Our cross-sectoral estimates suggest a link between declining dynamism and changes in characteristics.

Existing research has explored the reasons for declining entrepreneurship. Our findings add an important new piece of evidence: Even as the number of firms is declining, the types of entrepreneurs who start firms has changed.

1.1 Literature review

It is difficult to know at the time of founding whether or not firms are likely to survive and/or grow (Hathaway and Litan, 2014; Guzman and Stern, 2020). Indeed, the relationship between entrepreneurship and economic growth depends not simply on the quantity but also on the underlying quality of new firms (Schoar, 2010; Pugsley and Hurst, 2011). Prior attempts to use population-level data to characterize the rate of entrepreneurship have largely abstracted away from initial differences across entrepreneurs and focused on observable success. Notable exceptions are Guzman and Stern (2016); Fazio et al. (2016); Guzman and Stern (2020), who propose a measure of US startup quality based on observable startup characteristics (incorporation, firm name and organization, patent filing). Similar to Hombert et al. (2016), we use the French data to regress entrepreneur characteristics on observed success to determine which entrepreneur characteristics are predictive of success. We do so in order to measure startup quality over time, whereas Hombert et al. (2016)

do so to show that entrepreneur quality has not been affected by a 2002 reform of unemployment insurance for entrepreneurs.

Because young firms disproportionately contribute to employment and production (Davis and Haltiwanger 1992; Davis, Haltiwanger, and Shuh 1996; Haltiwanger, Jarmin, and Miranda 2013), the decline in the number of these firms is worrisome (Decker et al., 2016a,b). We contribute to the literature on entrepreneurship decline by showing similar facts on France. Our findings have implications for the debate on whether "responsiveness to shocks" or "shock dispersion" can explain the decline in entrepreneurship (Decker et al., 2018; Pugsley, Sedlacek, and Sterk, 2019). We focus on changes in entrepreneur characteristics to show that the *composition* of entrepreneurs has tilted towards more highly skilled entrepreneurs over time. In a way, this fact makes the decline in entrepreneurship more puzzling. Consistent with the existing literature, our results can be rationalized by an increased dispersion of shocks for startups ("the bar is getting higher"). We show that the long-term increase in startup quality is driven by sectors that have grown less.

The empirical literature has documented that start-ups' success depends on several ex ante characteristics of the founder and the firm more broadly (Pugsley, Sedlacek, and Sterk, 2019). Entrepreneurs' personal traits, risk aversion, and overconfidence levels to explain entrepreneurial entry and financial decisions at young firms (Moskowitz and Vissing-Jørgensen, 2002; Landier and Thesmar, 2008; Puri and Robinson, 2013; Kaplan, Klebanov, and Sorensen, 2012; Hvide and Panos, 2014; Levine and Rubinstein, 2017; Kerr, Kerr, and Xu, 2017). Key founder's characteristics include education and schooling (Bates, 1990; Levine and Rubinstein, 2017; Queiró, 2018), age (Ouimet and Zarutskie, 2014; Azoulay et al., 2020), gender (Howell and Nanda, 2019; Raina, 2019; Hebert, 2020), risk tolerance (Hvide and Panos, 2014), industry experience (Azoulay et al., 2020; Cassar, 2014), experience from entrepreneurial activity (Gompers et al., 2010; Cassar, 2014; Lafontaine and Shaw, 2016), entrepreneurial peers (Lerner and Malmendier, 2013), optimism (Landier and Thesmar, 2008). Firm's initial conditions have also been shown to be related to start-ups' success, in particular financing constraints (Kerr and Nanda, 2009), Venture Capital and Angel financing (Brav and Gompers, 1997; Hellmann and Puri, 2000; Kortum and Lerner, 2001; Kaplan, Sensoy, and Strömberg, 2009; Samila and Sorenson, 2011; Kerr, Lerner, and Schoar, 2014), as well as initial employment size, initial wage and initial employment growth (Maksimovic, Phillips, and Yang, 2019). Our dataset contains information on most of these entrepreneur characteristics, and our approach allows us to select those variables that are the most relevant in predicting success in order to construct a measure of startup quality over the years.

2 Data and summary statistics

2.1 Data sources

We use four sources of French administrative data provided by the French Statistical Office (INSEE): a survey of entrepreneurs conducted every four years, the exhaustive firm registry, accounting data from the tax files, and employment data from employer payrolls. Firms are uniquely identified by a 9-digit code (SIREN) that allows us to merge the different databases together.

2.1.1 Entrepreneur data

Our first source is the Système d'Information des Nouvelles Entreprises (SINE), which is a large-scale survey of entrepreneurs in France conducted by the French Bureau of Statistics every four years, from 1998 to 2014.[3] The main two advantages of these data is that they are not subject to any selection biases commonly encountered in the literature and that we are able to observe a large set of startups' founder characteristics. Questionnaires are sent to approximately 25% of entrepreneurs who started or took over a business in France that year. The surveyed firms are randomly selected from the exhaustive firm registry. The business owner is responsible for completing the documents. The response rate to the SINE survey is high (approximately 90%) because the tax authorities supervise the sending of questionnaires.

We focus on entrepreneurs who create a new startup by filtering out those who takeover an existing business (through purchase or inheritance, for instance). We exclude startups in the financial, agricultural, and public sectors from the sample. We obtain a representative sample of 175,366 new firms from the survey cohorts of 1998, 2002, 2006, 2010, and 2014. A few years after their inception, firms are resent similar questionnaires but we only focus on the initial survey. This survey contains information on the entrepreneur's main sociodemographic characteristics, experience, the reasons and motivations for which the firm was started, the conditions under which it was started (financing, initial research, customer prospects) and the founder's growth expectations. Table 1 contains summary statistics for some of these variables.

We explain in section 3 our approach to predict startup success based on entrepreneur characteristics from the SINE data. Our preferred prediction method restricts the set of SINE variables to entrepreneur characteristics that are *ex ante* characteristics in nature and do not consist in choices made by entrepreneurs upon the creation of their startup. We describe here this restricted set of

[3]These data are also used in Landier and Thesmar (2008); Hombert et al. (2020); Hebert (2020). The data are available through INSEE (click here).

variables and refer to appendix A for complete list of 48 entrepreneur characteristics available in the SINE data. We create a variable *Female* equal to one if the entrepreneur is a female. *AgeFounder* is the entrepreneur's age. *SameActivityBefore* is equal to one if the entrepreneur creates a firm in her previous sector of activity. *hasEntrepreneurRelatives* is a dummy equal to one if the entrepreneur has relatives who are entrepreneurs themselves, and *hasAnotherActivity* is equal to one if the entrepreneur simultaneously pursues another activity. Most variables are categorical, hence we encode them into dummies equal to one for each of the possible categories of a given variable. For instance, *Nationality* can take three values ("From the EU," "From France," and "From Outside the European Union") that result in three dummy variables equal to one if our observation of *Nationality* coincides with the dummy's category. Similarly, we build dummy variables corresponding to the observed categories for the entrepreneur's previous employment status (*NoActivity*, *UnemployedLessThanOneYear*, *UnemployedMoreThanOneYear*, *Employed*) and occupation (*BlueCollarOrCraftman*, *Employee*, *Executive*, *Independent*, *Unemployed*, etc.), the number of firms the entrepreneur created previous to the observed startup (*NoCreation*, *1Creation*, *2Creations*, *More3Creations*), the entrepreneur's education (*NoDegree*, *BelowHighSchool*, *HighSchool*, *HighSchool+2/3*, *Graduate*), the entrepreneur's previous employer size in number of employees (*NoPreviousEmployer*, *PrevEmployerLess10*, *PrevEmployer10-50*, *PrevEmployerMore50*).

[Insert Table 1 here]

Table 1 contains summary statistics on the *ex ante* entrepreneur characteristics we select. Because most of our variables are dummies, the reported means stand for percentage in the category. The only exception is age (AgeFounder). In our sample, 60% of the entrepreneurs create a startup in a sector they have previously worked in. Most entrepreneurs are exclusively working for their startups (80%) and are creating a startup for the first time (70%). Only 20% of the entrepreneurs are female, and the average age in our sample is 39 years old. 90% of the entrepreneurs are French and the remaining 10% are almost all from outside the European Union. 60% of entrepreneurs were working before creating a startup, whereas 20% were unemployed for less than a year, 10% were unemployed for more than a year, and the remaining 10% had no activity and no unemployment benefits. Entrepreneurs come from various backgrounds: blue collar workers, employees, executive, and unemployment are equally widespread occupations among entrepreneurs before they create a startup (20%). The second-most widespread occupation are independent and intermediate (10%). 50% of previously-working entrepreneurs were hired in a small firm with less than 10 employees.

2.1.2 Firm registry

The firm registry (*SIRENE*) contains the universe of firms registered in France from 1998 to 2017. For each newly created firm, the registry contains the industry the firm operates in based on a four-digit classification system similar to the four-digit SIC. It also provides the firm's legal status (e.g., Sole Proprietorship, Limited Liability Corporation, Corporation), the official creation date and geographical location.

2.1.3 Accounting data

Accounting data (balance sheet and income statements) is extracted from the tax files used by the Ministry of Finance for corporate tax collection purposes. The accounting information is therefore available for all French firms, public or private, whose annual sales exceed €32,600 (€81,500 in retail and wholesale trade).[4] We retrieve total sales and value added from the tax files. Sectors are defined by the French classification of sectors (*Nomenclature des activités Francaises, NAF*). We define sectors at the 2-digit industry level, and end up with 24 different sectors.

To create a consistent industry code along our sample period, we identify firms that exist both before and after the 2001 change from NAF1 to NAF2 industry classification change to calculate the fraction of firms in each NAF1 sector that belong to each NAF2 sector. For this sample, we keep the NAF1 codes for the firms' entire existence. For firms that are newly created after the NAF2 switch, we use two methods create a panel of firms at the NAF1 level. When we calculate aggregate statistics, we use the calculated probabilities to allocate newly-created firms in each NAF2 sector to the corresponding NAF1 sector. For individual-level estimates, such as those using the SINE survey, we assign each firm the most likely NAF1 sector (ie, the sector with the highest probability).

2.1.4 Employer payrolls

We use the French matched employer-employee dataset (*Déclarations Annuelles des Données Sociales*, DADS) to observe firms' employment. All firms that employ at least one employee must file payroll taxes. We use the DADS data to identify firm survival so that startups that never have any employees do not "survive" even in their first year of existence.

[4]Small firms with annual sales below this threshold can opt out and choose a special micro-business tax regime (*micro-entreprise*). These firms are not growth oriented. Income falling into this category is taxed at the individual level, hence they do not appear in the corporate tax files (Aghion et al., 2017). We exclude from our sample firms in the financial, agricultural and public sectors as they use different accounting systems.

2.2 Measures of startup success

To determine which entrepreneur characteristics are predictive of success, we need to define a measure of startup success. Our approach to identify high-growth start-ups considers the actual outcome for the firm. Because we are agnostic on what defines startup success, we consider several startup outcomes. Measuring a variety of outcomes also has the advantage to tell us whether successful startups are successful on every dimension or only specific ones. Our data allow us to study startups' survival, employment, sales, value added, at different horizons. Consistent with the existing literature, we focus on these outcomes at ages 3, 5, and 7.

Table 2 contains summary statistics of our success measures.

[Insert Table 2 here]

2.3 The evolution of startup dynamism in France

It is well documented that the startup entry rate has decreased in the US (see, e.g., Decker et al., 2014; Akcigit and Ates, 2021). Does a similar picture emerge in France? To investigate the evolution of the startup creation rate in France since the 1990s, we look at the evolution of the total number of new startups over the years and the share of employment from young firms.

Figure 1 plots the total number of new startups created every year from 1987 to 2019. Consistent with the evidence in Hombert et al. (2020), we find that the absolute number of startups increases around 2002. This number also has continuously increased from 2012 to 2019, when it reached its all-time peak. However, this evolution of the absolute number of startups can be misleading: In Appendix A1, we show that most of the new startups are sole proprietorships. In line with the growing literature arguing that self-employment is a poor proxy of entrepreneurship,[5] we focus on the number of startups created with at least one employee. The picture emerging from Figure 1 is now drastically different. As in the US, we find that the number of startups has steadily been declining since 1989. To the best of our knowledge, we are the first ones to document the decline in entrepreneurship in France.

[Insert Figure 1 here]

[5]Pugsley and Hurst (2011) document that most self-employed workers have no intention to grow or innovate. Haltiwanger, Jarmin, and Miranda (2013) show that after controlling for firm age, small businesses do not create jobs. Henrekson and Sanandaji (2014) find that the rate of self-made billionaires correlates negatively with self-employment rates. Schoar (2010) discusses the need to differentiate subsistence and transformational entrepreneurs. Levine and Rubinstein (2017) argue that incorporation is a better proxy of US entrepreneurship than self-employment.

A consequence of the declining startup rate is that the share of young firms in the economy, and the share of activity for which they account, is declining. Figure 2 shows that while 30% of firms were aged three years or less in 1994, this fraction fell to about 20% by 2015. This long-term continuous decline in the fraction of startups in the economy translated into a decrease in the share of employment accounted for by these young firms. This share fell from almost 15% in 1994 to around 5% in 2015.

[Insert Figure 2 here]

3 Measuring predicted startup quality

3.1 Which entrepreneur characteristics predict startup success?

In this section, we use machine learning techniques to determine which entrepreneur characteristics best predict startup success. Because the SINE data are highly dimensional (it contains up to 48 different variables on entrepreneurs and their startup at inception) and we don't know the true model of which entrepreneur characteristics predict their startup success, machine-learning techniques allow us to come up with better predictions than standard econometric techniques while remaining agnostic on the true underlying model mapping entrepreneur characteristics into startup success. The general model is:

$$Y_{i\tau} = \beta X_i + \varepsilon_{i\tau},$$

where i refers to one given entrepreneur/startup, $Y_{i\tau}$ is a measure of startup success (employment, value added, survival) at age τ (age 3, 5, or 7), and $\varepsilon_{i\tau}$ is an error term. We have up to 48 entrepreneur characteristics, hence the dimension of β is up to 48.

The LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996) approach is best suited for our analysis because it implements the model selection by selecting a restricted set of variables that best predict success.[6] We alleviate the attenuation bias induced by the penalized regression method of the LASSO by employing the Post-LASSO estimator that applies OLS to the variables selected by the first-stage variable selection method.[7] The idea is to establish correlations between startup success and *ex ante* entrepreneur characteristics. We are not trying to argue that

[6] In untabulated results, we use a standard OLS estimation, the ELASTIC NET technique, and the RIDGE selection model. Our predictive measures remain very similar across techniques.

[7] The post-estimation OLS writes

$$\hat{\beta}_{post} = \arg\min \frac{1}{N} \sum_{i=1} N(Y_{i\tau} - \beta X_i)^2 \text{ s.t. } \beta_j = 0 \text{ if } \tilde{\beta}_j = 0,$$

these characteristics are causally driving startup success at the exclusion of other variables, but we believe that they might be an indicator of some underlying fundamental difference of successful startups.

We try various specifications, including all 48 SINE variables and allowing for their interaction.[8] Our preferred specification focuses on 11 entrepreneur characteristics that reflect *ex ante* characteristics rather than choices of entrepreneurs (see section 2.2). In contrast to the existing literature (e.g., Guzman and Stern, 2020), this gives us the unique ability to measure changes in the *composition* of entrepreneurs and see if changes in predicted success are due to changes in the composition of entrepreneurs or in the time-varying success rate of the same types of entrepreneurs. A time-varying success rate of observationally equivalent entrepreneurs over time could be due to changes in the business environment or changes in the decisions made by the same types of entrepreneurs over time.

We train our LASSO algorithm on the 1998 cohort to compare the expected success of the subsequent cohorts of entrepreneurs irrespective of their actual observed success, relative to the 1998 cohort. Table 3 contains the results of the predictive LASSO regressions of our measures of success on the 11 ex ante entrepreneur characteristics. Although the LASSO estimates cannot be causally interpreted, we can check that the selected entrepreneur characteristics are correlated with startup success in a way that is expected by theory. For instance, across all specifications we find that whether entrepreneurs create a startup in a sector in which they have worked before is highly predictive of success. In line with the existing literature, we find that entrepreneur sex and age are also important predictors of startup success (see, e.g., Azoulay et al., 2020).

Interestingly, different entrepreneur characteristics are selected by the LASSO to predict firm survival on the one hand, and growth outcomes conditional on survival on the other. This means that a certain type of entrepreneurs are likely to survive for several years, whereas others are likely to grow large if they survive. For instance, in Table 3, entrepreneur age predicts survival, whereas previous executive experience predicts employment and value added growth. These observed differences across entrepreneur types validates the distinction made in the literature between "transformational" and "subsistence" entrepreneurs (Schoar, 2010).

[Insert Table 3 here]

where $\tilde{\beta}_j$ is a sparse LASSO estimator. Thus, the post-estimation OLS treats the first-step LASSO estimator as a genuine model selection technique.

[8] The results of other specifications for the LASSO will be added to our next draft.

To check the validity of our predictions, we plot in Figure 5 the mean realized success in the distribution of predicted success. The observed success of startups increases in the machine-learning predicted success rate based on entrepreneurs characteristics. These results suggest that machine learning models can be helpful in predicting whether a given entrepreneur's startup will be successful.

[Insert Figure 5 here]

4 The evolution of startup quality

4.1 Startup success and entrepreneur ability

In the previous section we showed that there is a decline in the importance of new firms for the French labor force. This raises the question of whether growth rates have changed across cohorts for the new firms that have been created.

One explanation for falling entrepreneurship is that it is becoming harder to start new firms. This could be because we are 'running out' of new ideas, for example, or because greater market power protects incumbents at the expense of new firms. On the one hand, if starting new firms is harder, it is natural to think that the average growth rates of new firms has fallen. On the other hand, if higher barriers causes low-quality entrepreneurs to give up altogether, then falling numbers of new firms could coincide with higher growth rates among the firms that do enter.

To explore these hypotheses, Figure 3 plots success rates over time for each cohort of firms included in the SINE survey. Each panel uses a different measure of success and tracks the average success rates over time for firms in each cohort.

The upper-left panel of Figure 3 plots average survival rates by cohort over time. Note that survival is measured as firms that appear in DADS data, so startups that never have any employees do not "survive" even in their first year of existence. Among the youngest firms, there is not a clear pattern for how survival has changed over time: Among two year old firms, 2002 cohort is the least likely to have employees, whereas the 2010 cohort is the most likely. By age 5, there appears to be a negative trend across cohorts, with later cohorts being less likely to have employees. However, the overall pattern is not that strong. Similarly, trends in value-add are unclear.

The lower-left panel shows log employment conditional on survival across cohorts. Here, the pattern is somewhat more clear: Later cohorts appear to have lower log employment. Whether this will continue to hold for the 2010 and 2014 cohorts remains to be seen.

There are two reasons that cohort-level changes in firm success that come from changing economic conditions may be difficult to measure. One is that temporary economic shocks can be hard to distinguish from long-run trends. For example, there is a dip in value-added that occurs at age 8 for the 2006 cohort, age 12 for the 2002 cohort and age 16 for the 1998 cohort of firms. This dip in value added is probably not an age effect but rather reflects negative shocks that happened in 2014.

The second difficulty is that changes in ex post firm success reflect both changes in the economic environment and endogenous changes in the characteristics of new entrepreneurs. One of the goals of this paper is to understand how these changing characteristics affect the measurement of survival and success probabilities.

To understand the effects of success on survival probabilities, Figure 4 plots average survival, log employment and log value-added for five-year-old firms across cohorts. The lines in this figure come from a firm-level regression that estimates how success varies across cohorts. Specifically, we estimate

$$Y_{it} = Cohort_i + \beta X_i + \varepsilon_{it}$$

on all firms of age 5 in the sample. The coefficients of interest are the cohort fixed effects, $Cohort_i$, which are plotted in the figure. To estimate the coefficients shown in green, we do not include any controls. The coefficients shown in orange come from specifications that include all the SINE variables in the vector of controls X_i.

Comparing the green and orange lines, we can see that controls substantially change the patterns of growth over time. Survival does not change much, but log employment and log value-added look substantially different across cohorts: Once we add controls, value-added at age 5 appears much lower in the 2006 cohort and even lower in the 2010 cohort, compared to specifications without controls. The same is true for log employment. Adding controls, these figures do suggest an overall downward trend in growth rates at the firm level.

We do not think that even the extensive controls in the SINE dataset measure all the important entrepreneur characteristics affecting later growth. For that reason, we do not take our estimates with controls to be the "true" average value added and growth rates across cohorts. Instead, the point of this figure is to demonstrate that adding controls matters a lot, which implies that entrepreneur characteristics are changing in a way that is economically important.

Another thing this figure hints at is the *direction* that entrepreneur characteristics are changing. In both the value-added and employment panels, adding entrepreneur controls reduces average growth rates. This means that the "quality" of entrepreneurs is improving over time, in the sense that their characteristics increasingly include variables that predict success. This result motivates the findings we explore in the next sections.

4.2 Time series patterns in entrepreneur characteristics

In Figure 6, we plot the average predicted 5-year value added and employment (in log) of the startups from the different cohorts.

The upper-left panel of Figure 6 shows that the predicted five-year employment of new startups has increased over time, reaching its peak in 2014 when our sample ends. A similar pictures emerges from looking at other success measures such as employment, shown in the lower-left panel. When we predict success at other horizons, such as three years or seven years, the pattern does not change much. We also get similar results when using interaction terms to predict success in the LASSO step, which greatly increases the number of predictors (the comparison between the main and interacted figures is shown in Figure A2 of the Appendix.)

The right panel of Figure 6 illustrates how the increase in startup quality stems from changes in entrepreneur characteristics over the years. We plot the evolution in the fraction of entrepreneurs with a college degree over the years as an example of an entrepreneur characteristics that has changed over time, explaining how our measure of startup quality has increased over the years. Importantly, our approach allows us to distinguish between two explanations for the increase in startup quality. Compared to an approach where we would look at realized success instead of predicted quality, our approach enables us to show that the increase in startup quality is due to changes in the *composition* of entrepreneurs rather than changes in the *business environment* that would explain why similar entrepreneurs would be more successful over time.

Next, we ask whether the same pattern holds for predicted entrepreneur survival. The upper-right panel of Figure 6 shows cohort-level changes in our measure of predicted five-year firm survival. The pattern here is different than for the measures for firm growth: The overall trend is negative, with a large shift happening between the 2004 and 2010 cohorts.

Putting together the results for firm survival with the results for growth conditional on survival, our results imply that French entrepreneurs are increasingly drawn from a population that less less likely to survive, but more likely to grow conditional on survival.

Previously, we showed that a different set of characteristics predicts growth as predicts survival. As discussed by Schoar (2010) for the case of developing countries, “transformational” and “subsistence” entrepreneurs likely have different characteristics and respond different sets of policies and economic conditions. This is consistent with our findings, which are consistent with changing trends in these types of entrepreneurs.

These findings also imply that the decline in high-growth firms may be even greater than previous research has shown. If entrepreneurs are increasingly drawn from a pool of individuals who are likely to create high-growth firms, then estimates which do not control for the change in composition may underestimate how much harder it is to create such firms.

[Insert Figure 6 here]

4.3 A similar trend among US entrepreneurs

Due to data availability, we cannot replicate our analysis for the US. However, we are able to show that similar to what happened in France, the past few decades have seen a decrease in the fraction of entrepreneurs in the US population while the fraction of entrepreneurs with a college degree has gone up.

Using data from the Survey of Consumer Finances (SCF), we measure the fraction of new entrepreneurs in the US population and the fraction of US entrepreneurs who report owning a college degree.[9] Appendix Figure A3 suggests that a similar trend of decreasing number but increasing quality of startups is happening in the US. To test whether the increase in the fraction of educated entrepreneurs is explained by a demographic trend of increasing education levels in the US, we regress in Appendix Table A1 the fraction of entrepreneurs with a college degree on survey years. The coefficient estimates of Survey Year confirm that the fraction of entrepreneurs with a college degree has gone up over the years. When controlling for entrepreneurs' birth year linearly in column (2), or via birth decade fixed effects in column (3), we find that the increase in educated entrepreneurs also holds irrespective of entrepreneurs' age. Because the new entrepreneurs from the same generation become more and more educated over the years, our interpretation is that the

[9] We identify entrepreneurs as individuals who (i) newly report working as self-employed, consultant, contractor, or in partnership (question X4106), and (ii) started the job less than 3 years ago (X4115) – i.e., after the last survey. We identify college graduates as individuals who report an Associate degree or above as their highest degree. Because survey weights (X42001) are at the household level, we identify entrepreneurs at the household level. A household is a new entrepreneur if either the reference person or the spouse is identified as a new entrepreneur, and it holds a college degree if the new entrepreneur within the household owns a college degree. For households with two new entrepreneurs, we use the highest education level across household members.

demographic trends cannot fully explain the increase in the quality of entrepreneurs.

5 Sectoral changes in entrepreneur composition

Controlling for entrepreneur characteristics makes the decline in entrepreneurship documented in Figure 2 puzzling. Indeed, it might seem contradictory that the share of young firms in total activity is going down despite the fact that entrepreneurs are increasingly "types" whose firms are likely to become large. So far, we have not provided evidence that these trends are related to each other. Here, we show sector-level patterns that suggest a relationship between these variables.

Here we show evidence consistent with the idea that rising entrepreneur quality and a falling entrepreneur share are in fact linked with each other. In Table 4, we ask whether there is a correlation between the decline in entrepreneurship and the increase in startup quality in the cross section of sectors. To do this, we estimate long-differences regressions at the sector level using the following specification:

$$\Delta Quality_s = \Delta(\frac{SmallFirmEmp_s}{TotalEmp_s}) + \varepsilon_s$$

where "$Quality_s$" refers to the change in average entrepreneur ability measured between 1998 and 2014. The independent variable in our specifications share of total employment represented by young firms (ages 1-5), measured as a fraction of total employment in the sector. This variable is one way of measuring the fall in startup share at the industry level; our results are similar when we simply measure the change in the number of startups or the fraction of employment from new firms. Our estimates are weighted by the number of firms in each sector in 1998.

We find that the long-term increase in startup quality over our sample period is driven by those sectors that grew less. This is true using all three measures of entrepreneurial quality. This suggests a link between entrepreneur quality and the changing economic importance of new firms.

[Insert Table 4 here]

One possible link between a falling entrepreneur share and rising entrepreneurial ability is due to entrepreneurial selection. Specifically, in sectors where the bar to entrepreneurship has increased, only higher-ability individuals are willing to start a new firm. We would expect that growing industries attract even marginal entrepreneurs, whereas in industries where success is relatively easy, even low-quality entrepreneurs would find it worthwhile to enter. This would mean that there is a negative correlation between sector-level growth and average predicted success.

To evaluate this hypothesis, we relate value-added per employee to changes in the characteristics of entrepreneurs. We repeat the previous specifications, but make our independent variable the change in value-added per employee among older firms. This is a proxy for the general economic conditions in the industry. The specification we estimate is:

$$\Delta Quality_s = \Delta LogValueAdded_s + \varepsilon$$

To measure $LogValueAdded$ at the sector level, we calculate the log value added per employee for every firm between ages 6 and 10 in the years 1998 and 2014. Then we calculate the average of this variable at the industry level in each year. The independent variable in our sector-level specification is the change in this variable between those years. As before, our estimates are weighted by the number of firms in 1998.

The results are shown in Table 5. In Columns (1) and (2), we link changes in value-added per employee to changes in the number of new firms. Unsurprisingly, industries where incumbent firms are growing have more startup entry in relative terms.

Columns (3), (4) and (5) show the link between average value-added per employee and the characteristics of new entrepreneurs. As hypothesized, we find a negative relationship. In industries where value-added growth is greater, there are both more new firms, and a lower average level of entrepreneur ability. The estimate is statistically significant for predicted value added and predicted survival, but not for predicted employment growth.

[Insert Table 5 here]

We interpret these results as the fact that the long-run decline in the number of startups is due to the fact that the bar to entrepreneurship is getting higher. Our findings provide evidence of a change in the selection of entrepreneurs, with only the most skilled entrepreneurs creating a startup towards the end of our sample (for instance, more educated entrepreneurs). This change in the composition of entrepreneurs coupled with a decrease in the share of activity due to young firms suggests that despite the increase in startup quality, skilled entrepreneurs are less able to grow and contribute to economic activity.

6 Conclusion

This paper establishes a new series of facts about entrepreneur characteristics. First, we show that entrepreneurs' performance is predictable based on ex ante characteristics. Different performance measures are predicted by different sets of entrepreneur characteristics. For example, demographic variables are better able to predict firm survival, but work experience is more important for high growth conditional on survival.

By predicting success using data from the 1998 cohort of firms, we are able to create measures of predicted success that we apply to later cohorts of firms. We establish that entrepreneur characteristics are changing — over time, there is a decrease in characteristics that predict survival, and an increase in characteristics predictive of growth.

Finally, we show sectors with lower growth had a greater increase in entrepreneur ability over time. This finding links our results to the large literature on entrepreneurial dynamism. It suggests that the decline in the number of firms, and the increase in the fraction of high-growth entrepreneurs, may be linked. A possible reason for this is self-selection into entrepreneurship.

Research on entrepreneurship has been focused on trends in the declining number of startups. Our findings suggest a second moment for this research to explain — changes in the number of entrepreneurs.

Tables and Figures

Figure 1. Declining Number of Startups with at Least One Employee

Source: Firm creations. This figure plots the evolution over time of the number of startups in France (green line, left axis) and of the number of startups with at least one employee at creation (orange line, right axis).

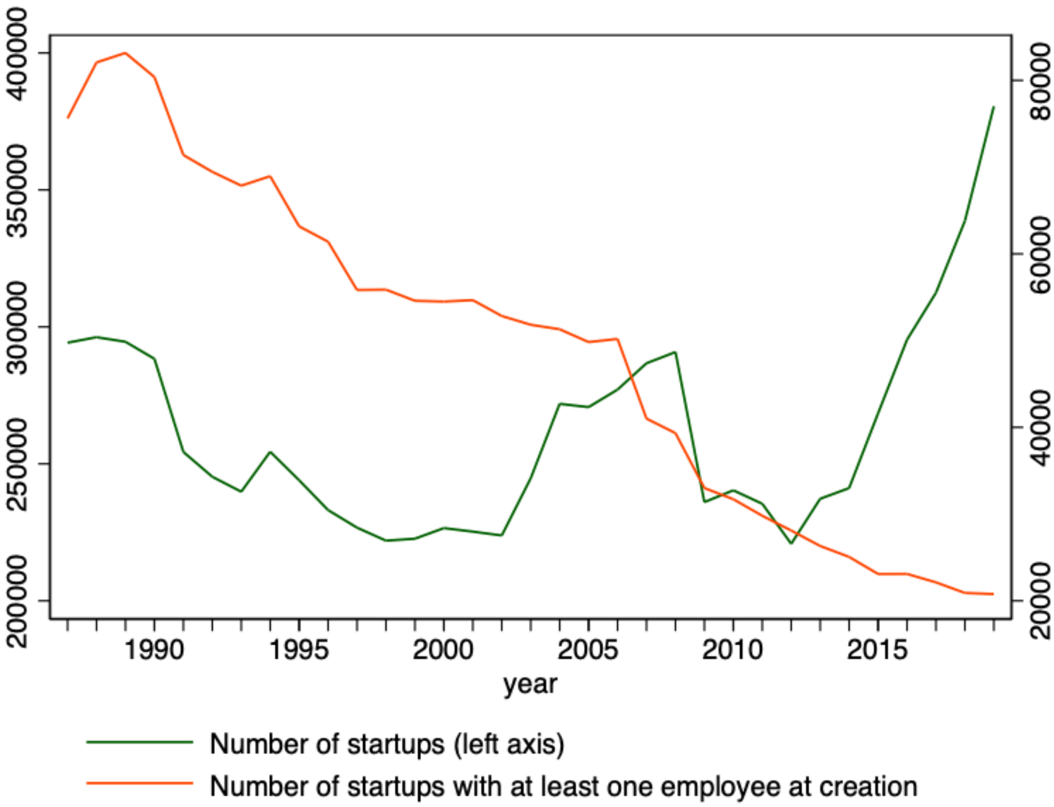

Figure 2. Declining Share of Employment from Young Firms (Firms Age 5 or Less)

Source: Matched employer-employee dataset (DADS). This figure plots the evolution over time of the fraction of total employment in firms with age of 1-3 years (green line) as well as the fraction of all firms with age of 1-3 years.

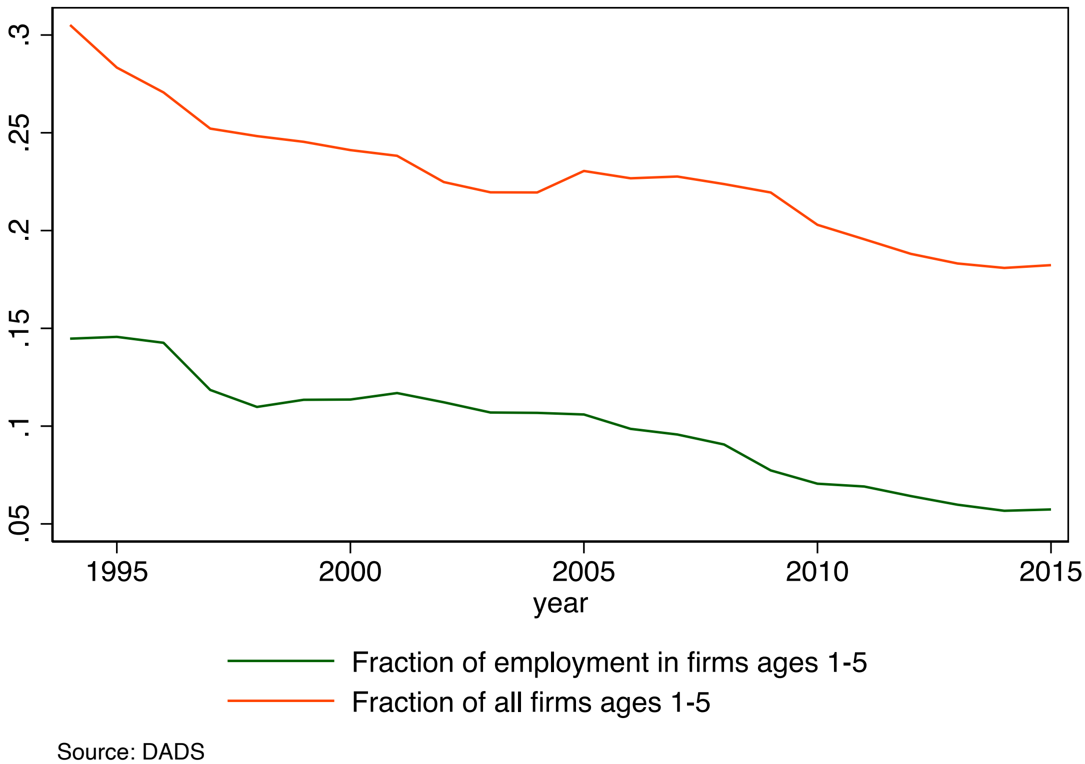

Figure 3. Success Rates Over Time, by Cohort

Source: SINE surveys (1998, 2002, 2006, 2010, 2014), Matched employer-employee dataset (DADS). This figure plots average success for each cohort of firms. The measures of realized success are log employment and log value-added (conditional on firm survival) and an indicator for firm survival. Survival is measured based on whether the firm reports any payroll information.

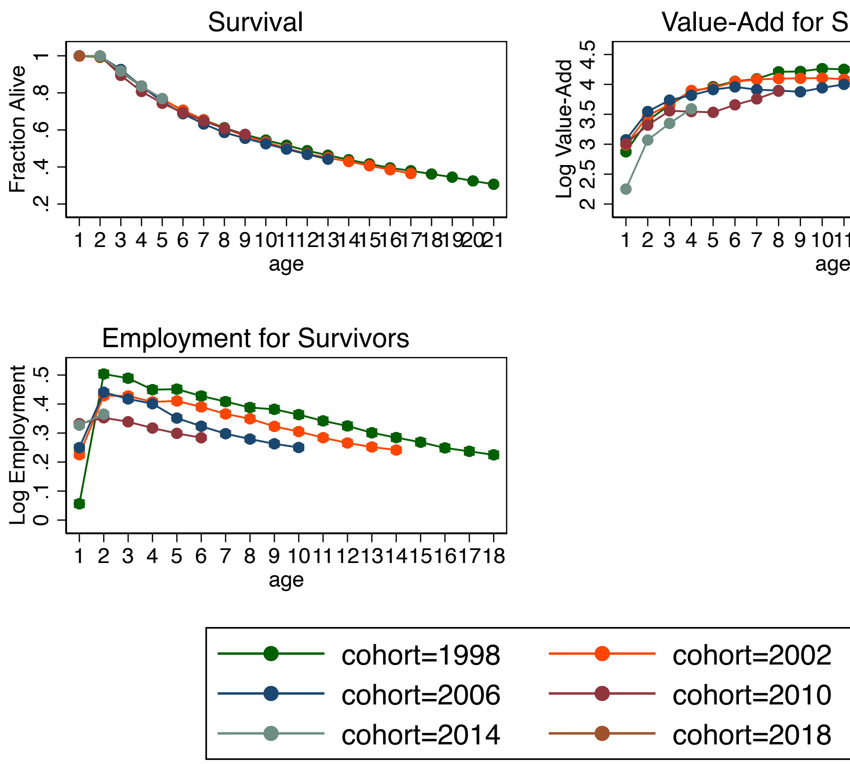

Figure 4. Effect of Controls on Success by cohort

Source: Matched employer-employee dataset (DADS). This figure plots average realized success for five-year old firms in each cohort of french startups in SINE. The measures of realized success are log employment and log value-added (conditional on firm survival) and an indicator for firm survival. Survival is measured based on whether the firm reports any payroll information. The measures with controls are the result of estimates from a linear regression of success on cohort indicators, which are graphed, and linear controls for other entrepreneur characteristics.

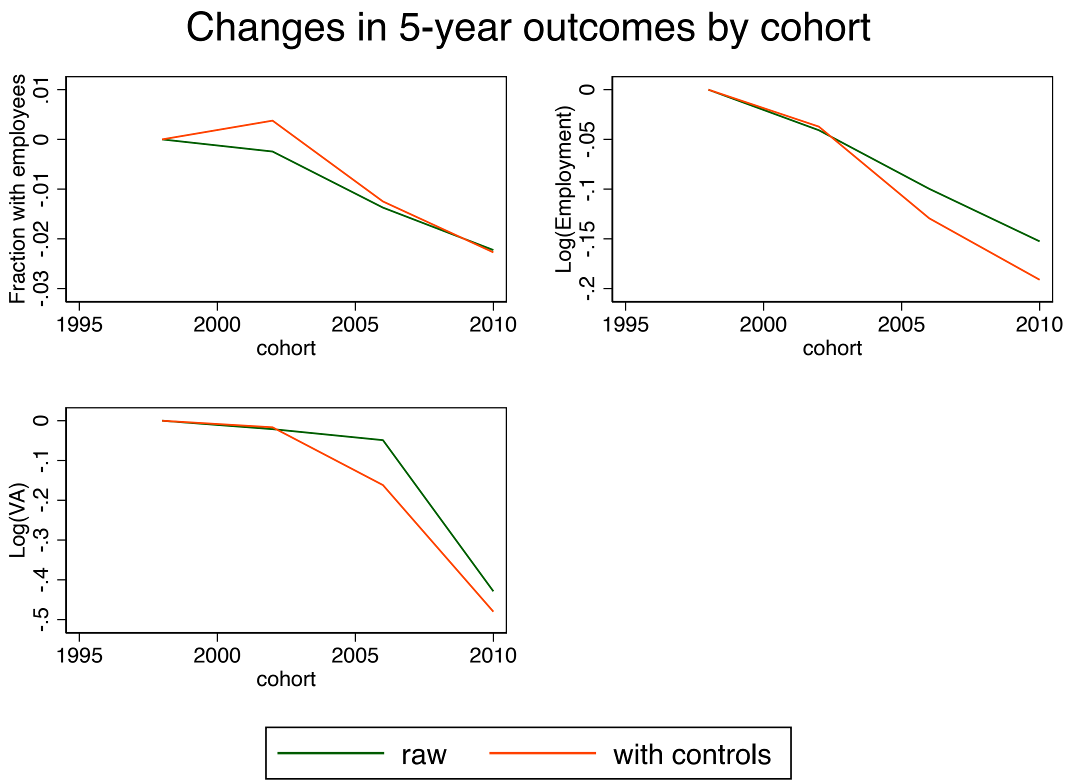

Figure 5. Startup Quality Measure: Predicted versus Realized Success

Source: SINE, tax files, matched employer-employee dataset (DADS). This figure plots realized firm-level success against deciles of predicted success. The measures of realized success are log employment and log value-added (conditional on firm survival) and an indicator for firm survival. Survival is measured based on whether the firm reports any payroll information.

(a) (b) (c)

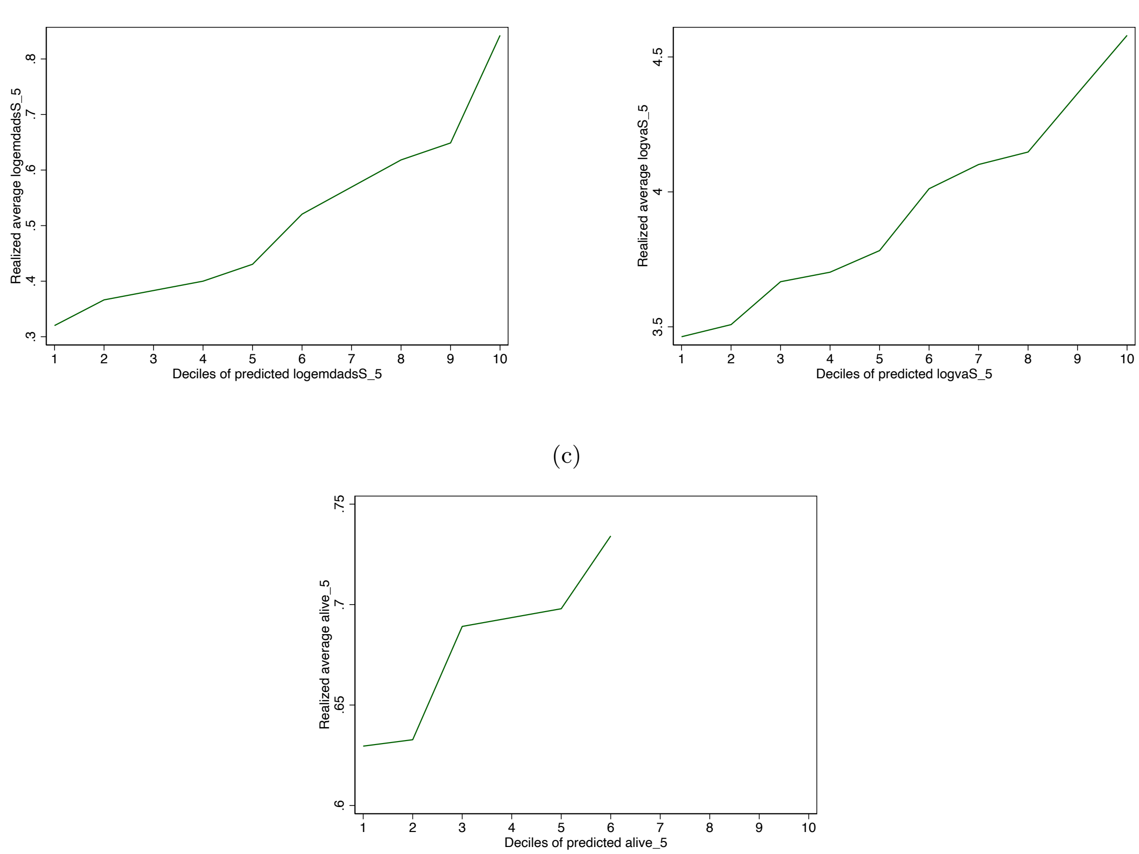

Figure 6. Increasing Startup Quality and Changes in Entrepreneur Characteristics

Source: SINE, tax files, matched employer-employee dataset (DADS). This figure plots changes in entrepreneur quality by cohort, measured as the post-LASSO estimates of predicted success.

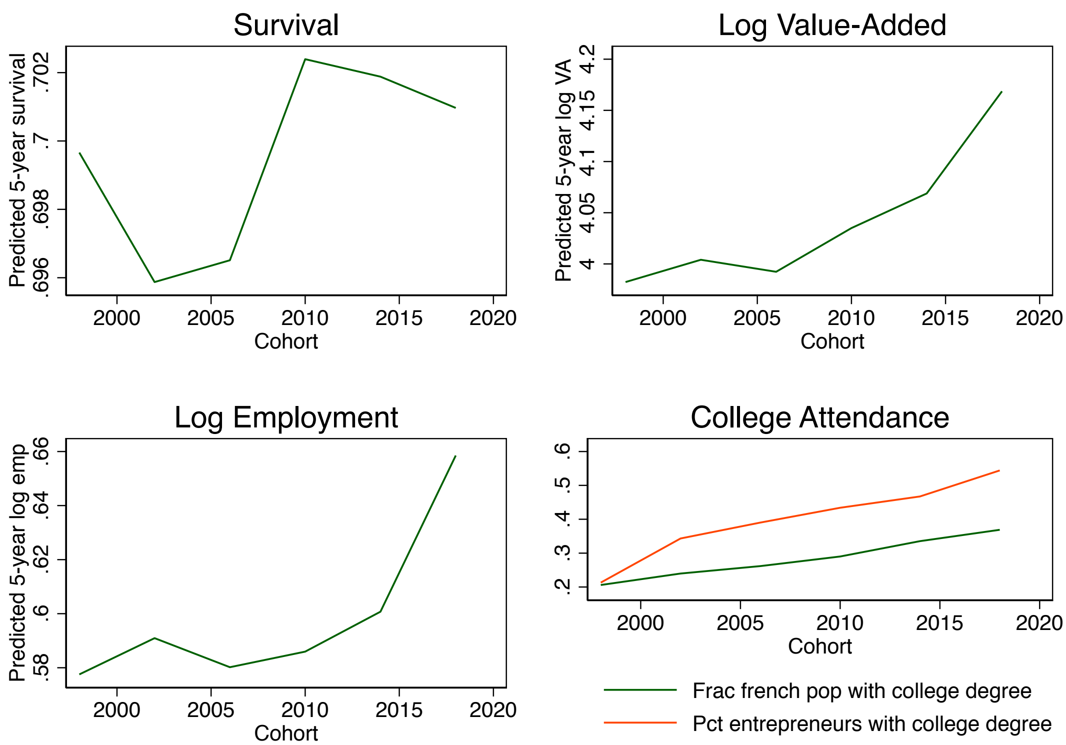

Table 1: **Summary statistics on entrepreneur characteristics**

Source: 1998, 2002, 2006, 2010, and 2014 SINE surveys. This table contains summary statistics for the main predictive variables selected in our LASSO procedure to predict startup success. Most variable are dummies, so that the reported means stand for percentage in the category. The only exception is age (AgeFounder).

	Obs	Mean	Sd	5%	25%	50%	75%	95%
SameActivityBefore	107,491	0.6	0.5	0.0	0.0	1.0	1.0	1.0
Female	107,865	0.2	0.4	0.0	0.0	0.0	0.0	1.0
hasEntrepreneurRelatives	107,505	0.7	0.5	0.0	0.0	1.0	1.0	1.0
hasAnotherActivity	107,940	0.2	0.4	0.0	0.0	0.0	0.0	1.0
AgeFounder	107,860	39.3	10.4	24.5	31.0	38.0	47.0	57.5
Nationality	.	.	.	.	.	.	.	.
FromEU	107,708	0.0	0.2	0.0	0.0	0.0	0.0	0.0
FromFR	107,708	0.9	0.3	0.0	1.0	1.0	1.0	1.0
FromOutsideEU	107,708	0.1	0.2	0.0	0.0	0.0	0.0	1.0
LastJobType	.	.	.	.	.	.	.	.
BlueCollarOrCraftman	107,518	0.1	0.3	0.0	0.0	0.0	0.0	1.0
Employee	107,518	0.2	0.4	0.0	0.0	0.0	0.0	1.0
Executive	107,518	0.2	0.4	0.0	0.0	0.0	0.0	1.0
Independent	107,518	0.1	0.3	0.0	0.0	0.0	0.0	1.0
Intermediate	107,518	0.1	0.3	0.0	0.0	0.0	0.0	1.0
NoJob	107,518	0.2	0.4	0.0	0.0	0.0	0.0	1.0
SeveralJobs	107,518	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Student	107,518	0.0	0.2	0.0	0.0	0.0	0.0	0.0
PreviousSituation	.	.	.	.	.	.	.	.
NoActivity	107,511	0.1	0.3	0.0	0.0	0.0	0.0	1.0
UnemployedLess1Y	107,511	0.2	0.4	0.0	0.0	0.0	0.0	1.0
UnemployedMore1Y	107,511	0.1	0.4	0.0	0.0	0.0	0.0	1.0
Working	107,511	0.5	0.5	0.0	0.0	1.0	1.0	1.0
Education	.	.	.	.	.	.	.	.
Bac	106,367	0.1	0.3	0.0	0.0	0.0	0.0	1.0
Bac+2/3	106,367	0.2	0.4	0.0	0.0	0.0	0.0	1.0
Bac+5+	106,367	0.2	0.4	0.0	0.0	0.0	0.0	1.0
BelowBac	106,367	0.4	0.5	0.0	0.0	0.0	1.0	1.0
NoDegree	106,367	0.1	0.3	0.0	0.0	0.0	0.0	1.0
PrevEmployerSize	.	.	.	.	.	.	.	.
NoPreviousEmployer	98,166	0.1	0.2	0.0	0.0	0.0	0.0	1.0
PrevEmployer10-50	98,166	0.2	0.4	0.0	0.0	0.0	0.0	1.0
PrevEmployerLess10	98,166	0.5	0.5	0.0	0.0	0.0	1.0	1.0
PrevEmployerMore50	98,166	0.2	0.4	0.0	0.0	0.0	0.0	1.0
NbFirmCreation	.	.	.	.	.	.	.	.
1Creation	107,501	0.2	0.4	0.0	0.0	0.0	0.0	1.0
2Creations	107,501	0.1	0.2	0.0	0.0	0.0	0.0	1.0
More3Creations	107,501	0.0	0.2	0.0	0.0	0.0	0.0	0.0
NoCreation	107,501	0.7	0.5	0.0	0.0	1.0	1.0	1.0

Table 2: **Summary statistics on startup success measures**

Source: Firm registry (SIRENE), tax files (BIC) and employer payrolls (DADS). This table reports summary statistics for the main variables used as startup success for firms that appear in the SINE entrepreneur survey.

	Obs	Mean	Sd	5%	25%	50%	75%	95%
$\mathbb{1}$(Alive at age 3)	108,799	0.8	0.4	0.0	1.0	1.0	1.0	1.0
$\mathbb{1}$(Alive at age 5)	85,041	0.7	0.5	0.0	0.0	1.0	1.0	1.0
$\mathbb{1}$(Alive at age 7)	85,041	0.6	0.5	0.0	0.0	1.0	1.0	1.0
Employment at age 3	87,324	0.3	0.7	0.0	0.0	0.0	0.0	1.8
Employment at age 5	60,660	0.4	0.8	0.0	0.0	0.0	0.7	2.1
Employment at age 7	48,174	0.3	0.7	0.0	0.0	0.0	0.0	1.9
Value Added at age 3	66,074	3.6	1.5	1.3	2.7	3.8	4.6	5.9
Value Added at age 5	43,596	3.8	1.5	1.1	2.9	3.9	4.7	6.1
Value Added at age 7	37,510	3.8	1.6	1.3	2.8	3.9	4.8	6.3

Table 3: **LASSO-selected variables for 5-year startup employment**

Sources: 1998 SINE survey cohort, tax files, matched employer-employee dataset (DADS), registry (SIRENE). We train our LASSO algorithm on the 1998 cohort only, to later compare the expected success of the subsequent cohorts of entrepreneurs irrespective of their actual observed success. Sectors refer to an industry at the 2-digit level of the French SIC. We discuss our dependent variables in section 2.2 and our approach in section 3. Standard errors are double-clustered at the sector of origin and sector of entry level and are reported in parentheses. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

	𝟙(Alive at age 5)			log(EmploymentAge5)			log(ValueAddedAge5)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SameActivityBefore	0.064***	0.065***	0.055***	0.118***	0.106***	0.083***	0.217***	0.201***	0.163***
	(0.011)	(0.011)	(0.009)	(0.023)	(0.022)	(0.022)	(0.037)	(0.036)	(0.034)
Female	-0.052***	-0.050***					-0.158***	-0.131***	
	(0.013)	(0.013)					(0.043)	(0.041)	
AgeFounder	0.004***	0.003***	0.002***						
	(0.001)	(0.001)	(0.001)						
FromOutsideEU	-0.230***	-0.197***	-0.172***						-0.340***
	(0.031)	(0.028)	(0.026)						(0.091)
NoJob	-0.110***	-0.088***	-0.092***						
	(0.026)	(0.024)	(0.022)						
Bac+5+	-0.067***	-0.060***							
	(0.015)	(0.014)							
NoCreation	0.055***	0.050***		-0.108***	-0.106***	-0.144***		-0.092**	-0.091**
	(0.014)	(0.014)		(0.029)	(0.029)	(0.028)		(0.043)	(0.041)
Executive				0.300***	0.271***		0.517***	0.429***	0.348***
				(0.036)	(0.035)		(0.044)	(0.046)	(0.045)
Working				0.175***	0.160***	0.140***	0.337***	0.290***	0.251***
				(0.022)	(0.021)	(0.020)	(0.029)	(0.029)	(0.026)
Intermediate						0.045			
						(0.037)			
PrevEmployer10-50							0.153***	0.131***	
							(0.033)	(0.032)	
FromFR								0.228***	
								(0.060)	
Constant	Yes	No	No	Yes	No	No	Yes	No	No
Zone FE	No	Yes	No	No	Yes	No	No	Yes	No
Zone $\times$ Industry FE	No	No	Yes	No	No	Yes	No	No	Yes
Observations	7,474	7,466	7,739	5,684	5,669	5,290	5,410	5,391	5,011
R^2	0.03	0.03	0.02	0.04	0.04	0.02	0.08	0.07	0.05

Table 4. Long-run Changes in Number of Startups, and Startup Quality at the Sector Level

Sources: SINE, tax files, DADS. The dependent variables are the long-term changes in quality across sectors and the independent variable is the long-term change in the share of total employment represented by young firm at the sector level. Sectors refer to an industry at the 2-digit level of the French SIC. Robust standard errors. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

	(1) Pred Log(Young VA)	(2) Pred Survival	(3) Pred Log(Employment)
Change in Workforce in Firms Ages 1-5	-0.43***	-0.055***	-0.073**
	(0.095)	(0.020)	(0.028)
Constant	0.056***	-0.021***	0.022***
	(0.011)	(0.0027)	(0.0039)
Observations	38	38	38
R^2	0.269	0.094	0.079

Table 5. Changes in Industry Value-Added and Startup Quality at the Sector Level

Sources: SINE, tax files, DADS. The dependent variables are the long-term changes in sector-level growth and the independent variable is the log value added per employee for every firm between ages 6 and 10 in the years 1998 and 2014.. Sectors refer to an industry at the 2-digit level of the French SIC. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively.

	(1) Log(Young Firms)	(2) Log(New Firms)	(3) Avg Predicted VA	(4) Avg Predicted Emp	(5) Avg Predicted Survival
Chg in Log VA per Emp	2.289**	2.022*	-0.212*	-0.0563	-0.0552**
	(2.74)	(2.38)	(-2.20)	(-1.66)	(-3.02)
Constant	-0.815***	-0.760***	0.148***	0.0427***	-0.00396
	(-4.06)	(-3.73)	(5.91)	(4.86)	(-0.94)
Observations	35	34	34	34	34
R^2	0.252	0.207	0.131	0.079	0.157

Bibliography

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva. 2017. "Tax simplicity and heterogeneous learning." Tech. rep., National Bureau of Economic Research.

Akcigit, Ufuk and Sina T Ates. 2021. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." *American Economic Journal: Macroeconomics* 13 (1):257–98.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda. 2020. "Age and high-growth entrepreneurship." *American Economic Review: Insights* 2 (1):65–82.

Bates, Timothy. 1990. "Entrepreneur human capital inputs and small business longevity." *The review of Economics and Statistics* :551–559.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014a. "High-dimensional methods and inference on structural and treatment effects." *Journal of Economic Perspectives* 28 (2):29–50.

———. 2014b. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies* 81 (2):608–650.

Brav, Alon and Paul A Gompers. 1997. "Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and nonventure capital-backed companies." *The journal of finance* 52 (5):1791–1821.

Cassar, Gavin. 2014. "Industry and startup experience on entrepreneur forecast performance in new firms." *Journal of Business Venturing* 29 (1):137–151.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The role of entrepreneurship in US job creation and economic dynamism." *Journal of Economic Perspectives* 28 (3):3–24.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda. 2016a. "Declining business dynamism: What we know and the way forward." *American Economic Review* 106 (5):203–07.

———. 2016b. "Where has all the skewness gone? The decline in high-growth (young) firms in the US." *European Economic Review* 86:4–23.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda. 2018. "Changing business dynamism and productivity: Shocks vs. responsiveness." Tech. rep., National Bureau of Economic Research.

Fazio, Catherine, Jorge Guzman, Fiona Murray, and Scott Stern. 2016. "A new view of the skew: Quantitative assessment of the quality of American entrepreneurship." *Kauffman Foundation New Entrepreneurial Growth, Kansas City, MO: February* .

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein. 2010. "Performance persistence in entrepreneurship." *Journal of financial economics* 96 (1):18–32.

Guzman, Jorge and Scott Stern. 2016. "The State of American Entrepreneurship: New Estimates of the Quality and Quantity of Entrepreneurship for 32 US States, 1988-2014." Tech. rep., National Bureau of Economic Research.

———. 2020. "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." *American Economic Journal: Economic Policy* 12 (4):212–43.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda. 2013. "Who creates jobs? Small versus large versus young." *Review of Economics and Statistics* 95 (2):347–361.

Hathaway, Ian and Robert Litan. 2014. "Declining business dynamism: It's for real." *Economic Studies at Brookings, Brookings Institution. https://www. brookings. edu/research/declining-business-dynamism-its-for-real* .

Hebert, Camille. 2020. "Mind the Gap: Gender Stereotypes and Entrepreneur Financing." .

Hellmann, Thomas and Manju Puri. 2000. "The interaction between product market and financing strategy: The role of venture capital." *The review of financial studies* 13 (4):959–984.

Henrekson, Magnus and Tino Sanandaji. 2014. "Small business activity does not measure entrepreneurship." *Proceedings of the National Academy of Sciences* 111 (5):1760–1765.

Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar. 2016. "Does Unemployment Insurance Change the Selection into Entrepreneurship?" In *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*. University of Chicago Press, 351–369.

———. 2020. "Can unemployment insurance spur entrepreneurial activity? Evidence from France." *The Journal of Finance* 75 (3):1247–1285.

Howell, Sabrina T and Ramana Nanda. 2019. "Networking frictions in venture capital, and the gender gap in entrepreneurship." Tech. rep., National Bureau of Economic Research.

Hvide, Hans K and Georgios A Panos. 2014. "Risk tolerance and entrepreneurship." *Journal of Financial Economics* 111 (1):200–223.

Kaplan, Steven N, Mark M Klebanov, and Morten Sorensen. 2012. "Which CEO characteristics and abilities matter?" *The Journal of Finance* 67 (3):973–1007.

Kaplan, Steven N, Berk A Sensoy, and Per Strömberg. 2009. "Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies." *The Journal of Finance* 64 (1):75–115.

Kerr, Sari Pekkala, William Kerr, and Tina Xu. 2017. "Personality traits of entrepreneurs: a review of recent literature." .

Kerr, William and Ramana Nanda. 2009. "Financing constraints and entrepreneurship." Tech. rep., National Bureau of Economic Research.

Kerr, William R, Josh Lerner, and Antoinette Schoar. 2014. "The consequences of entrepreneurial finance: Evidence from angel financings." *The Review of Financial Studies* 27 (1):20–55.

Kortum, Samuel and Josh Lerner. 2001. *Does venture capital spur innovation?* Emerald Group Publishing Limited.

Lafontaine, Francine and Kathryn Shaw. 2016. "Serial entrepreneurship: Learning by doing?" *Journal of Labor Economics* 34 (S2):S217–S254.

Landier, Augustin and David Thesmar. 2008. "Financial contracting with optimistic entrepreneurs." *The Review of Financial Studies* 22 (1):117–150.

Lerner, Josh and Ulrike Malmendier. 2013. "With a little help from my (random) friends: Success and failure in post-business school entrepreneurship." *The Review of Financial Studies* 26 (10):2411–2452.

Levine, Ross and Yona Rubinstein. 2017. "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics* 132 (2):963–1018.

Maksimovic, Vojislav, Gordon M Phillips, and Liu Yang. 2019. "Do Public Firms Respond to Industry Opportunities More Than Private Firms? The Impact of Initial Firm Quality." Tech. rep., National Bureau of Economic Research.

Moskowitz, Tobias J and Annette Vissing-Jørgensen. 2002. "The returns to entrepreneurial investment: A private equity premium puzzle?" *American Economic Review* 92 (4):745–778.

Ouimet, Paige and Rebecca Zarutskie. 2014. "Who works for startups? The relation between firm age, employee age, and growth." *Journal of financial Economics* 112 (3):386–407.

Pugsley, Benjamin W, Petr Sedlacek, and Vincent Sterk. 2019. "The nature of firm growth." *Available at SSRN 3086640* .

Pugsley, Benjamin Wild and Eric Hurst. 2011. "What do small businesses do?" *Brookings Papers on Economic Activity* 43 (2):73–142.

Puri, Manju and David T Robinson. 2013. "The economic psychology of entrepreneurship and family business." *Journal of Economics & Management Strategy* 22 (2):423–444.

Queiró, Francisco. 2018. "Entrepreneurial human capital and firm dynamics." *Available at SSRN 3280925* .

Raina, Sahil. 2019. "VCs, founders, and the performance gender gap." *Finance Down Under 2017 Building on the Best from the Cellars of Finance* .

Samila, Sampsa and Olav Sorenson. 2011. "Venture capital, entrepreneurship, and economic growth." *The Review of Economics and Statistics* 93 (1):338–349.

Schoar, Antoinette. 2010. "The divide between subsistence and transformational entrepreneurship." *Innovation policy and the economy* 10 (1):57–81.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1):267–288.

Appendix

A SINE variables

We describe below the list of all entrepreneur characteristics available in the SINE data. We encode categorical variables into dummies equal to one for each of the possible categories of a given variable.

Continuous variables:

- $AgeFounder$: the entrepreneur's age
- $NbDirectorsNonSalaried$: the new firm's number of non salaried directors
- $NbDirectorsSalaried$: the new firm's number of salaried directors
- $NbWorkersTotal$: the new firm's total number of workers

Dummy variables:

- $Female$: dummy equal to one if the entrepreneur is a female
- $SameActivityBefore$: dummy equal to one if the entrepreneur creates a firm in her previous sector of activity
- $hasEntrepreneurRelatives$: dummy equal to one if the entrepreneur has relatives who are entrepreneurs themselves
- $hasAnotherActivity$: dummy equal to one if the entrepreneur simultaneously pursues another activity
- $isSubsidiary$: dummy equal to one if the new firm is a subsidiary
- $FinancedBankLoan$: dummy equal to one if the new firm is financed by a bank loan
- $FinancedOtherLoan$: dummy equal to one if the new firm is financed by another type of loan
- $FinancedPersonalResources$: dummy equal to one if the new firm is financed by the entrepreneur's personal resources
- $FinancedOtherFirmVC$: dummy equal to one if the new firm is financed by another firm of a Venture Capitalist
- $FinancedSubsidy$: dummy equal to one if the new firm is financed by subsidies
- $PublicSubsidy$: dummy equal to one if the new firm obtained a public subsidy
- $MotivatedNewIdea$: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was a new idea
- $MotivatedIndependence$: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was more independence

- *MotivatedOpportunity*: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was a specific opportunity
- *MotivatedEntrepreneurPeer*: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was the presence of entrepreneur peers
- *MotivatedNoJobDecidedToStart*: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was that she had no job but decided to start the firm
- *MotivatedNoJobForcedToStart*: dummy equal to one if the entrepreneur declared that her motivation to start the new firm was that she had no job and was forced to start the firm
- *Franchised*: dummy equal to one if the new firm is a franchise
- *UseExternalServiceAccounting*: dummy equal to one if the new firm uses external accounting services
- *UseExternalServiceMgmt*: dummy equal to one if the new firm uses external management services
- *UseExternalServiceLogistic*: dummy equal to one if the new firm uses external logistic services
- *UseExternalServiceCleaning*: dummy equal to one if the new firm uses external cleaning services
- *UseExternalServiceSale*: dummy equal to one if the new firm uses external sale services
- *UseExternalServiceOther*: dummy equal to one if the new firm uses another type of external services
- *WantToGrow*: dummy equal to one if the entrepreneurs declared she wanted to grow the firm in the coming years
- *WantToMaintain*: dummy equal to one if the entrepreneurs declared she wanted to maintain the firm's business in the coming years
- *WantToRecover*: dummy equal to one if the entrepreneurs declared she wanted to recover in the coming years
- *WillHire*: dummy equal to one if the entrepreneurs declared she wanted to hire people in the coming years

Categorical variables:

- *Nationality*: categorical variable corresponding to the founder's nationality (*FromEuropeanUnion*, *FromFrance*, and *FromOutsideEuropeanUnion*)
- *PreviousSituation*: categorical variable corresponding to the entrepreneur's previous employment status (*NoActivity*, *UnemployedLess1Y*, *UnemployedMore1Y*, *Working*)
- *LastJobType*: categorical variable corresponding to the observed categories for the entrepreneur's previous occupation (*BlueCollarOrCraftman*, *Employee*, *Executive*, *Independent*, *Unemployed*, *Intermediate*, *NoJob*, *SeveralJobs*, *Student*)

- $PrevEmployerSize$: categorical variable corresponding to the entrepreneur's previous employer size in number of employees (*NoPreviousEmployer*, *PrevEmployerLess10*, *PrevEmployer10-50*, *PrevEmployerMore50*)
- *Education*: categorical variable corresponding to the entrepreneur's education (*NoDegree*, *BelowHighSchool*, *HighSchool*, *HighSchool+2/3*, *Graduate*)
- $NbFirmCreation$: categorical variable corresponding to the number of firms the entrepreneur created previous to the observed startup (*NoCreation*, *1Creation*, *2Creations*, *More3Creations*)
- $nonFinancialHelpType$: categorical variable corresponding to the non financial help the entrepreneur received (*Agency*, *Alone*, *Family*, *Professional*)
- $InitialCapital$: categorical variable corresponding to the new firm's initial capital in Euro (*Less2k*, *2k-4k*, *4k-8k*, *8k-16k*, *16k-40k*, *40k-80k*, *+80k*)
- $CompanyMgmtType$: categorical variable corresponding to the new firm's management type (*Alone*, *Associates*, *Family*, *Spouse*)
- $Innovation$: categorical variable corresponding to whether the new firm is innovative (*NewMethodOrOrganization*, *NewProductOrServiceOrSale*, *No*)
- *NbClient*: categorical variable corresponding to the new firm's number of clients (*1-2Client*, *3-10Clients*, *10+Clients*)
- *TypeClient*: categorical variable corresponding to the new firm's type of clients (*Firms*, *Retail*, *Public*)
- *LocationClient*: categorical variable corresponding to the location of the new firm's clients (*Local*, *Regional*, *National*, *International*)

B Additional figures and tables

Figure A1. Declining Number of Startups with at Least One Employee

Source: Firm registry. This figure plots the evolution in the number of newly-created startup in different legal forms (limited liability, simplified shares, etc.). We restrict the plot to the most common legal forms.

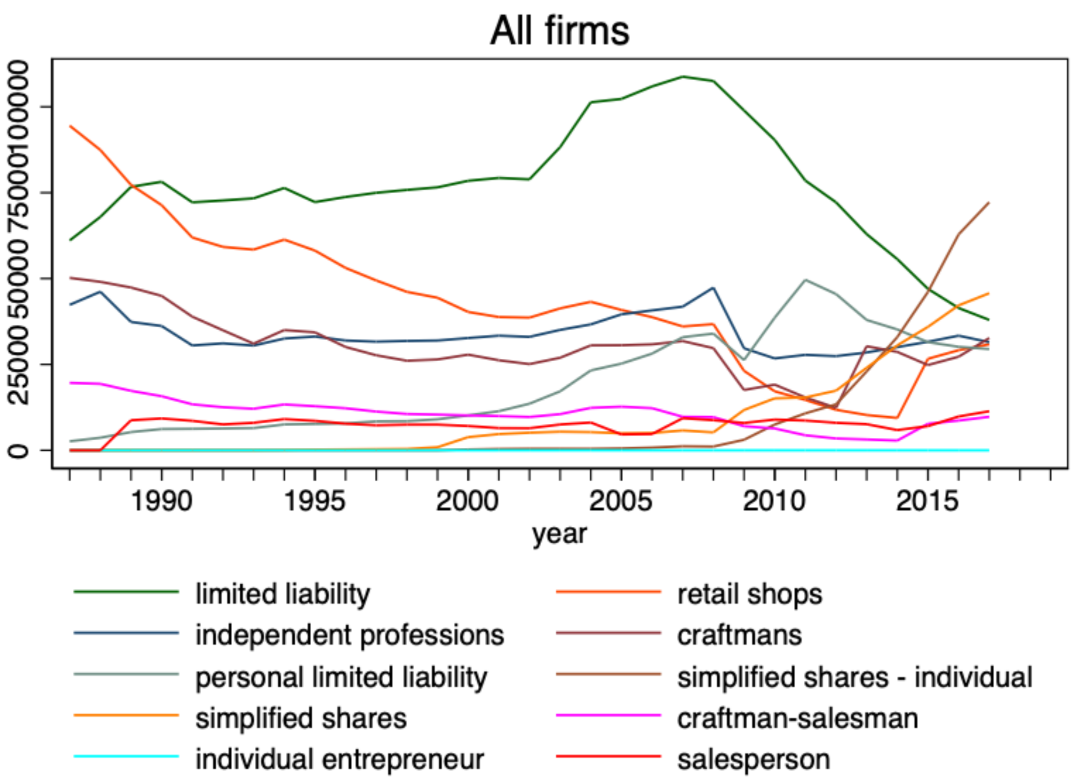

Figure A2. Increasing Startup Quality and Changes in Entrepreneur Characteristics: The Role of Interacted Variables in our Quality Measure

Source: SINE, tax files, matched employer-employee dataset (DADS). This figure plots the post-LASSO estimates of predicted startup outcomes at age 5 (log value added, log employment, and survival) over the five SINE cohorts of entrepreneurs (1998, 2002, 2006, 2010, and 2014). We show the predictions of the simple LASSO and of the LASSO that allows for interactions among dependent variables.

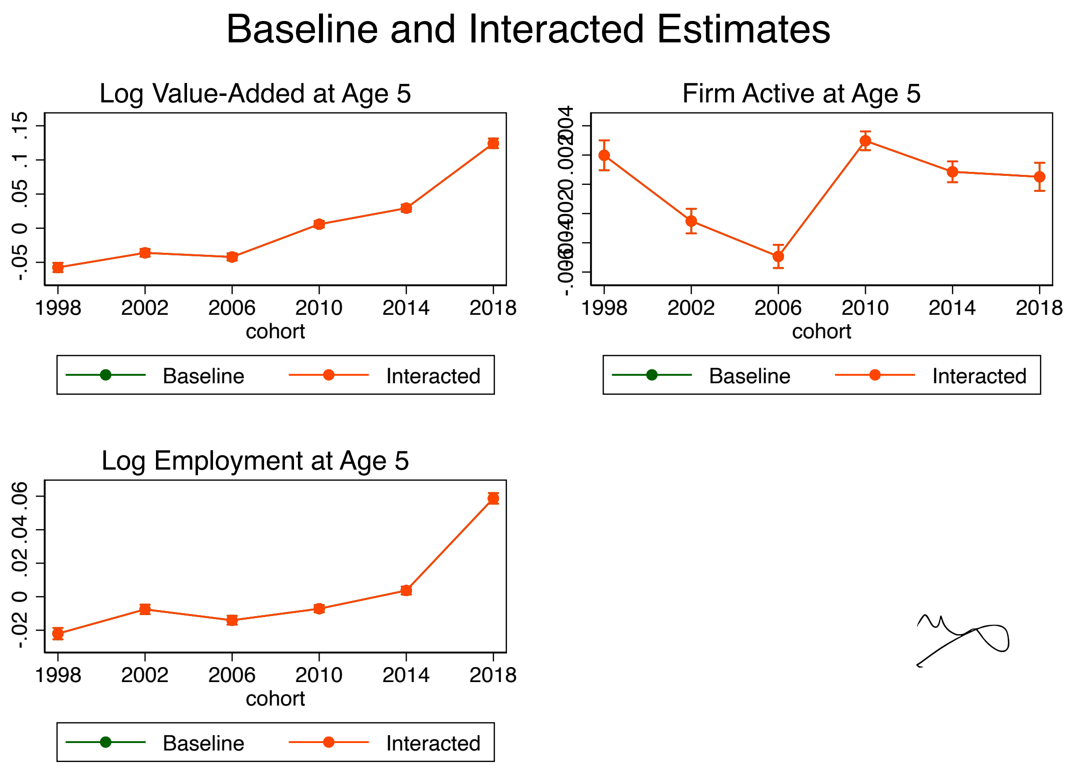

Figure A3. Declining fraction of entrepreneurs and increasing entrepreneur quality: Suggestive evidence on US data

Source: Survey of Consumer Finances (SCF). This figure plots the fraction of the US population that reports "entrepreneur" as an occupation and the fraction of entrepreneurs who report owning a college degree. We identify entrepreneurs at the household level, when either the reference person or the spouse newly reports working as self-employed, consultant, contractor, or in partnership at the time of the survey and this person has started the job less than 3 years before the survey year (i.e., between two surveys). If the household has at least one entrepreneur according to that definition, the dummy variable *College Degree* is equal to one if any entrepreneur in the household holds an associate college degree or above. The household variables are aggregated to each survey year using the SCF replication weights.

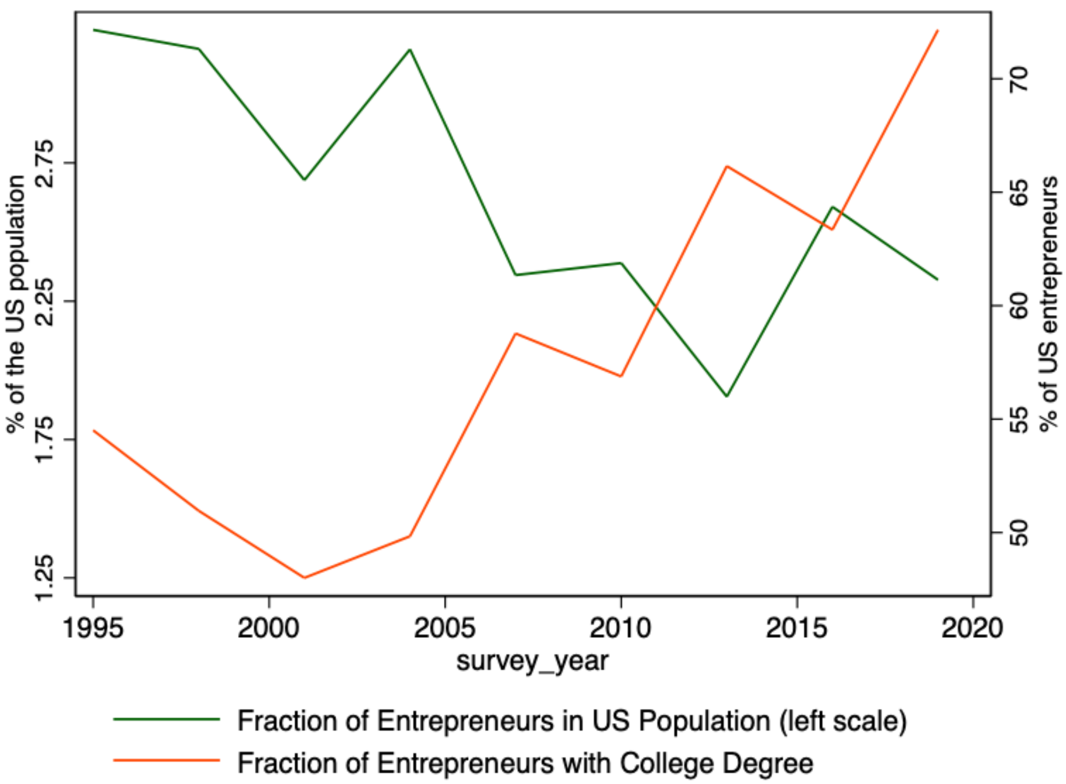

Table A1. Increasing entrepreneur education versus demographic changes: Suggestive evidence on US data

Source: Survey of Consumer Finances (SCF). This table shows that the fraction of US entrepreneurs who report owning a college degree has increased more than the fraction of the US population owning a college degree over our sample period (1995–2019). We identify entrepreneurs at the household level, when either the reference person or the spouse newly reports working as self-employed, consultant, contractor, or in partnership at the time of the survey and this person has started the job less than 3 years before the survey year (i.e., between two surveys). If the household has at least one entrepreneur according to that definition, the dummy variable *College Degree* is equal to one if any entrepreneur in the household holds associate college degree or above. The household variables are aggregated to each survey year using the SCF replication weights. *, **, and *** denote results that are significantly different from zero at the 10, 5 and 1% levels, respectively. Standard errors are clustered at the household level.

	College Degree	College Degree	College Degree
Survey Year	0.0082***	0.012***	0.013***
	(0.0020)	(0.0022)	(0.0022)
Entrepreneur Birth Year		-0.0048***	
		(0.0013)	
Constant	-15.9***	-14.4***	-24.7***
	(4.02)	(4.08)	(4.49)
Observations	7129	7129	7129
Birth Decade FE	No	No	Yes
Adjusted R2	0.017	0.032	0.036